

科学数据政策研究的热点及演化趋势分析☆

宋永辉[1,2,3]　姜鑫[2]　马廷灿[1,3*]
1 中国科学院大学经济与管理学院图书情报与档案管理系　北京 100190
2 黑龙江大学信息管理学院　哈尔滨 150080
3 中国科学院武汉文献情报中心　武汉 430000
* 通信作者E-mail: matc@whlib.ac.cn

[摘　要] 当前，科学数据管理工作受到世界各国科研机构的高度关注，对比分析国际与国内科学数据政策研究领域的异同点有利于推进该领域研究的发展。选取科学数据政策相关论文作为研究对象，使用文献计量与内容分析法分析国际与国内该领域研究发展特征与演化趋势，揭示该领域的研究热点。研究结果表明：国际、国内科学数据政策研究均处于探索发展阶段，研究文献量呈波动上升趋势；研究方向以信息科学与图书馆学为主导，同时部分研究属于交叉学科成果；国际研究较多关注科研数据共享，国内侧重科学数据政策、数据管理及科研数据间的研究关系。
[关键词] 科学数据政策　演化趋势　研究热点　文献计量　内容分析
DOI: 10.15978/j.cnki.1673-5668.202106001

1　引言

近年来，随着互联网、大数据、人工智能等信息技术的迅猛发展，科技创新工作愈发依赖于科学数据的获取和分析，科学数据存储与共享的重要性日益凸显[1]，科学数据政策的制定与实施受到了国际各国政府组织和科研机构的高度重视[2]。目前，在国际层面通过制定相关政策以保障和规范科学数据的合理、有效管理与共享已成为全球共识[3,4]。2010年，澳大利亚国家数据服务中心发布的《高校/机构科研数据管理政策大纲》[5]为澳洲高校制定科研数据政策提供指导，有效地促进了澳大利亚高校科研数据政策的发展。2011年，英国研究理事会发布《RCUK数据政策共同原则》，提出科研数据的开放和共享将会促进科学研究的发展，总结了科研数据管理与共享的原则要求[6]。为加速研究、支持新发现与促进合作，澳大利亚、英国、美国等国家均已制定出涵盖国家、政府职能部门、科研机构、科研资助机构、高等院校、数据存储/共享中心及学术出版集团等不同层面

☆本研究受到国家社会科学基金项目“科学数据开放政策分析与评估研究”（项目编号：16BTQ060）、2020年中国科学院文献情报能力建设专项“科技领域战略情报研究咨询体系建设”（项目编号：E0290001）与中国科学院2020年科创计划基金（项目编号：20204001841 ）的资助与支持。

的科学数据政策体系[7,8]。我国科研群体也对科学数据管理与共享工作形成了较为统一的共识[9]。为进一步规范我国科学数据的管理与共享工作，国务院办公厅于2018年3月发布了《科学数据管理办法》（以下简称“《办法》”）[10]。《办法》指出，国务院和省级政府应负责建立健全本部门（本地区）的科学数据管理政策，科研院所和高等院校及企业等单位应负责建立科学数据管理系统。此外，中国科学院也于2019年2月发布《中国科学院科学数据管理与开放共享办法（试行）》[11]，提出以明确主体责任、明确工作机制、规范业务流程以及规范适用范围等四个要素为核心的科学数据管理基本框架。综上，合理构建科研数据制度已势在必行，制定和实施科研机构层面的科研数据政策将成为我国科研机构的核心任务之一[12]。

基于此，本文旨在利用文献计量与内容分析的方法，揭示国际与国内科学数据政策研究领域的发展现状、研究热点与研究主题演化趋势，并重点探讨以下几个问题：（1）国际、国内科学数据政策研究领域的发展状况如何；（2）国际、国内该领域不同时期的研究主题与研究热点是什么；（3）国际、国内在该领域的研究工作存在哪些异同点。

2 数据来源与研究方法

2.1 数据来源

鉴于发表于北大中文核心期刊、CSSCI来源期刊与*Web of Science*核心合集数据库的期刊论文具有较高的学术认可度，能够较好地反映出具体研究领域的研究动态和发展趋势[13]，因此本文以中国知网CNKI数据库中北大中文核心期刊、CSSCI来源期刊与*Web of Science*核心合集SSCI、SCI-E数据库收录的科学数据政策研究文献为数据源。

在中国知网CNKI数据库中检索中文研究文献，具体检索策略如下：主题=（科学数据政策）OR（科研数据政策），时间跨度=（2000－2021年），来源期刊类型选择“北大核心”与“CSSCI”，删除与本文研究主题不相关的文献，并通过参考文献与施引文献进行补充，共获得126篇中文研究文献。

在*Web of Science*核心合集的SSCI和SCIE数据库中检索英文研究文献，具体检索策略如下：TS=（“scientific data polic*”）OR TS=（“research data polic*”），时间跨度=（2000－2021年），通过人工判读删除与研究主题不相关的文献，并通过参考文献与施引文献进行补充，共获得80篇外文研究文献。数据采集时间为2021年4月27日。

2.2 研究方法

本研究主要采用文献计量与内容分析方法，同时辅以文献调研与网络调查法。首先详细阅读和整理分析国际、国内相关文献，进行简要评述，综述科学数据政策研究发展现状；其次，通过网络检索引擎从政府、科研机构、科研资助机构及高等院校等层面获取国际、国内科学数据政策文本与通知文件等相关资料；再次，采用文献计量法探索国际、国内该领域的发展历程与特征、主导研究学科领域及研究热点；最后，在国际、国内研究文献数量特征分析的基础上，通过内容分析法挖掘国际、国内科学数据政策研究不同阶段的重点研究主题，揭示该领域研究的演进趋势。本文用到的统计分析与可视化工具有Excel、NoteEXpress、Ucinet、SPSS 21.0等。

3 结果分析

3.1 发文量分布趋势分析

3.1.1 国际发文年代分析

年度发文量与变化趋势可以反映该领域的总体发展特征。整体而言，国际层面科学数据政策研究领域发文量经历了从平缓发展到迅速增加再到波动上升的变化过程（参见图1）。依据文献的发表时间，可将国际科学数据政策研究分为3个发展阶段：萌芽期（2007－2013年），增长期（2014－2016年），波动期（2017－2020年）。在萌芽阶段，相关研究文献增幅较小。2013年底，美国“棱镜计划”“棱镜门”事件曝光，世界各国



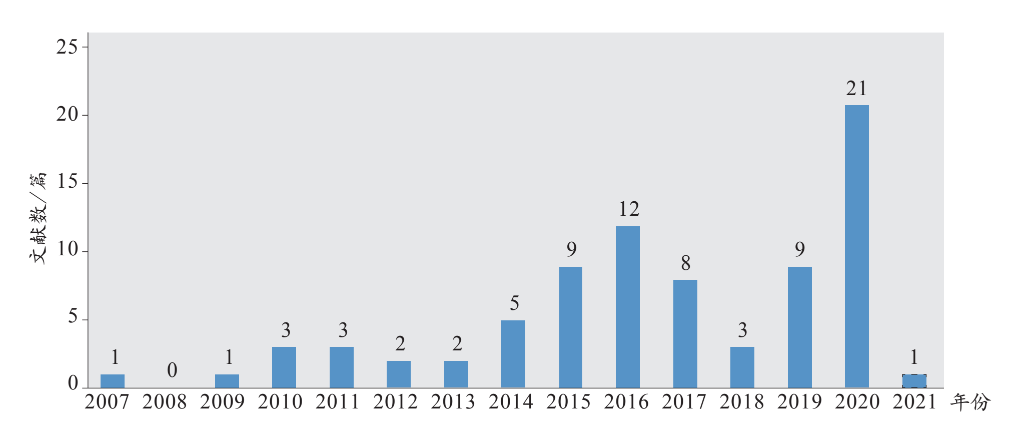


图1　国际科学数据政策研究领域发文量年度分布图

注：图中2021年数据的虚线边框表示数据不完整。

开始高度重视数据安全[14]。2014－2016年该领域文献数量增长较快，相关研究成果显著增加。2017－2020年，相关发文量呈波动上升趋势。总体来看，国际科学数据政策研究成果呈现萌芽发展到快速增加再到波动增长的趋势，预测该领域未来将受到更多学者关注，属于当前热点研究领域[15]。

3.1.2　国内发文年代分析

可以看出，国内科学数据政策研究领域最早的文献发表于2008年，2013年起发文量迅速增多（参见图2）。

根据上述年发文数的变化可将国内科学数据政策文献研究分为2个发展阶段：萌芽期（2008－2010年）和增长期（2013－2020年）。萌芽期的文献数量由0篇涨为6篇，增幅较小，相关研究产出较少。2013年开始进入增长期，文献增长速度明显提高，特别是2016年以来年度发文数量快速提升。这在一定程度上说明在当前大数据、深度学习与人工智能兴起的同时，国内学者愈来愈重视科学数据政策研究。总体来看，国内该领域年发文量具有明显的持续增长趋势，可以预测未来几年该领域的研究发展将持续向好。

综上，国际、国内研究人员对于科学数据政策研究领域的研究兴趣方兴未艾，未来可能迎来科学数据政策研究的黄金时代，科学数据管理、开放及共享的原理与技术将比今天有更广泛和更深入的应用。

3.2　发文期刊分析

3.2.1　国际发文量TOP10期刊

为进一步了解国际科学数据政策文献来源期刊，

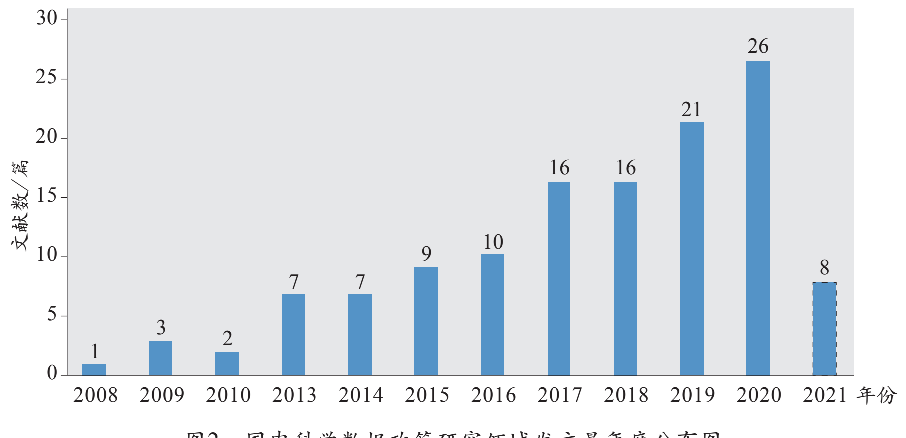


图2　国内科学数据政策研究领域发文量年度分布图

注：图中2021年数据的虚线边框表示数据不完整。

本节根据各期刊的文章发表量，得到国际科学数据政策研究发文量TOP10期刊（参见表1）。由表可知，*Scientometrics*刊载8篇研究成果，是发文量最多的期刊，其次是*PLoS One*与*Journal of the Association for Information Science and Technology*，载文量分别为5篇和4篇。载文量4篇及以上的3种期刊中，*Scientometrics*是情报学与图书馆学领域的重要期刊；*PLoS One*是多学科科学领域的重要期刊；*Journal of the Association for Information Science and Technology*是信息科学与计算机科学领域的重要期刊。刊登在此类国际高水平期刊上的文章代表了当前国际科学数据政策研究领域的研究热点与重点。

整体来看，相关研究较多集中于信息科学与图书馆学领域专业期刊，但也有少数成果发表于计算机科学、公共卫生与环境卫生、物理学与应用经济学等领域的顶级专业期刊，表明科学数据政策研究涉及领域较为广泛，且正在受到来自除信息科学与图书馆学外的更多学科领域学者的关注和研究。

3.2.2 国内发文量TOP10期刊

国内科学数据政策研究发文量TOP10期刊如表2所示。《图书情报工作》发文量最多，为12篇。其次是《情报杂志》《情报理论与实践》《数字图书馆论坛》与《图书情报知识》，分别为11、10、7、6篇。论文刊登数量5篇及以上的7种期刊中，《图书情报工作》《情报杂志》《图书情报知识》《情报理论与实践》及《图书馆论坛》的复合影响因子均在2.5以上，属于国内图书情报与数字图书馆领域高水平刊物，这类期刊上的论文基本代表了国内科学数据政策研究领域的研究热点与重点。

虽然上述发文较多的期刊主要为图书情报与数字图书馆的专业期刊，但还有诸如《中国科技资源导刊》《中国科学院院刊》及《地球科学进展》等载文量介于1～3篇的其他学科重要期刊。可见国内科学数据政策研究涉及领域同样较为广泛。

综上，国际、国内科学数据政策研究领域文献的来源出版物的学科领域分布结果比较一致，说明科学数据政策方面的研究热点处于以信息科学与图书馆学为主导的研究方向。此外，国际、国内该领域研究也发表于其他学科重要期刊，其研究成果能够代表各自领域与科学数据政策研究领域知识交叉的研究，由此可见科学数据政策文献部分研究成果属于交叉学科。

表1 国际科学数据政策研究领域刊载量TOP10期刊

序号	期刊名称	载文量/篇
1	*Scientometrics*	8
2	*PLoS One*	5
3	*Journal of the Association for Information Science and Technology*	4
4	*Insights-the UKSG Journal*	3
5	*Peer J*	3
6	*Malaysian Journal of Library Information Science*	2
7	*Science Editing*	2
8	*Science and Engineering Ethics*	2
9	*European Journal of Human Genetics*	2
10	*Library Management*	2

表2 国内科学数据政策研究领域刊载量TOP10期刊

序号	期刊名称	载文量/篇
1	图书情报工作	12
2	情报杂志	11
3	情报理论与实践	10
4	数字图书馆论坛	7
5	图书情报知识	6
6	图书馆论坛	6
7	图书馆学研究	5
8	情报资料工作	5
9	图书馆建设	4
10	中国科技资源导刊	4

3.3 研究热点分析

3.3.1 国际关键词词频分析

关键词是对文章核心的高度概括和精炼，通过对高频关键词进行分析，可以进一步归纳总结科学数据政策研究领域的研究趋势与当前热点[16]。为深入分析国际科学数据政策研究领域的研究热点，本节对目标文献的关键词进行统一处理，删除不规范的关键词，如现状、对策、分析等无明确意义的关键词，而后提取国际科学数据政策文献高频关键词得出该领域高频关键词表（参见表3）。

统计结果表明，“data sharing（数据共享）”“open data（开放数据）”“research data（科学数据）”“research data management（科学数据管理）”是国际该领域研究中出现频次较高的关键词。由此可见国际该领域研究主要集中于科学数据的管理、开放及共享。

3.3.2 国内关键词词频分析

同样提取国内科学数据政策文献高频关键词得出该领域前10位高频关键词表（参见表4）。

经统计发现，国内科学数据政策研究领域的关注热点是科学数据（53）、数据共享（23）、数据政策（21）、科研数据（19）等。此外，科研数据政策、科研数据共享等新兴关键词也呈现较高的频次变化率，说明科研数据政策、科研数据共享及开放数据[17]的研究空间较大，是该领域未来的潜在研究热点。

国际、国内科学数据政策研究在科学数据的政策、开放与共享等热点主题领域交叉，属于国际、国内共同关注的研究重点。

表3 国际科学数据政策研究领域前10位高频关键词及其出现频次

序号	关键词	频次
1	data sharing	24
2	open data	11
3	research data	6
4	research data management	6
5	collaboration	5
6	data management	4
7	reproducibility	4
8	research integrity	3
9	Africa	3
10	scholarly publishing	2

表4 国内科学数据政策研究领域前10位高频关键词及其出现频次

序号	关键词	频次
1	科学数据	53
2	数据共享	23
3	数据政策	21
4	科研数据	19
5	数据管理	17
6	科研数据管理	15
7	科学数据管理	14
8	开放共享	12
9	政策	11
10	科学数据共享	10

3.4 高频关键词共现分析

3.4.1 国际高频关键词共现分析

共词分析（Co-word Analysis）的思想来源于文献计量学的引文耦合与共被引概念，当两个关键词同时出现在一篇文献中时，则称这两个关键词存在共现关系（co-occurrence）[18]。本节按上述理论，利用NoteExpress软件提取得到科学数据政策研究文献的高频关键词共现矩阵（Co-word Matrix），而后通过社会网络分析软件Ucinet、NetDraw展现各关键词间的共现强弱关系（参见图3）。热点关键词共现图谱中节点大小表示该关键词同其他关键词共现频次，频次越高，节点越大；两节点之间连线表示两个关键词同时出现在同一篇文献中，连线越粗，表示两关键词共现次数越多，关系越密切[19,20]。

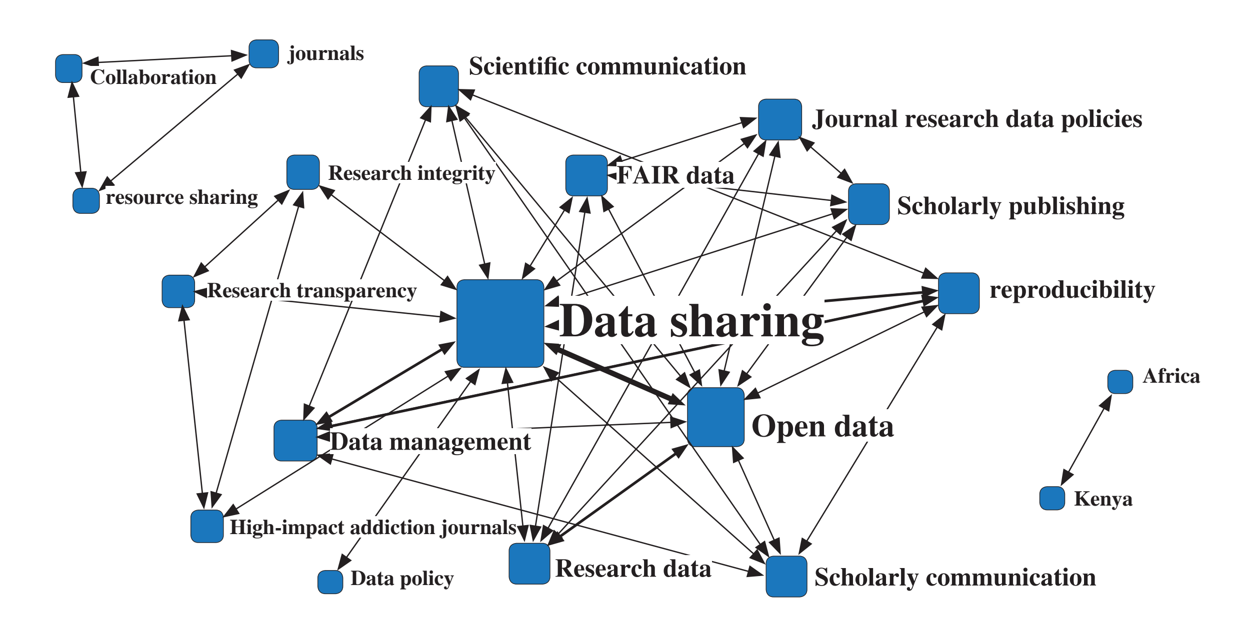


图3　国际科学数据政策研究领域高频关键词分布图谱（词频≥2）

由图可知，国际科学数据政策研究的关注热点是“data sharing（数据共享）”“open data（开放数据）”“research data management（科学数据管理）”“collaboration（合作）”“research data（科研数据）”等。此外，“data sharing（数据共享）”和“open data（开放数据）”以及“data sharing（数据共享）”和“data management（数据管理）”间的连线较粗，对应两个关键词共现次数较多，表明国际学者侧重于科学数据的管理、开放与共享研究，这与关键词词频统计的分析结论相照应。

3.4.2　国内高频关键词共现分析

同样利用上述方法得到国内科学数据政策研究领域高频关键词分布图谱（参见图4）。由图可知，节点较大的关键词有“数据政策”“开放数据”“科研数据管理”“数据管理”“科研数据共享”“科研数据政策”等，一定程度上代表了国内科学数据政策研究领域的研究热点。此外，“科学数据”与“数据共享”“数据管理”“数据政策”间连线较粗，表明“科学数据”与这些关键词之间共现次数较多，反映国内学者较多开展科学数据管理与共享政策的研究。

3.5　高频关键词聚类分析

3.5.1　国际高频关键词聚类分析

本节使用系统聚类法对科学数据政策研究领域的高频关键词展开分析。将初始共词矩阵通过Salton指数法转换为相关矩阵（Correlation Matrix）[19]。Salton指数法的计算公式为：$S = C_{ij} / (C_i \times C_j)^{\frac{1}{2}}$，其中$C_i$、$C_j$分别代表关键词 i 与关键词 j 的词频，C_{ij}为关键词与关键词的共现次数，通过统计分析软件SPSS 21.0对高频关键词进行系统聚类，得到国际该领域研究中高频关键词系统聚类树状图（参见图5）。

国际科学数据政策研究高频关键词的聚类群体划分情况如表5和图5所示，上述高频关键词可分成7个聚类，分别代表国际科学数据政策研究领域的7个重要研究主题：（1）聚类1为科学数据管理，侧重于科学数据二次利用等管理办法；（2）聚类2为科研数据开放与共享原则，主要包括探讨科技期刊与科研机构的研究数据开放与共享原则；（3）聚类3为科研诚信与科研透明度，侧重于对科研诚信和科研数据（或成果）透明度研究；（4）聚类4为期刊数据共享，侧重于科技期刊的研究数



据共享研究；（5）聚类5为欠发达国家科研数据管理研究，主要对发展中国家的科研数据管理现状与未来发展策略进行分析；（6）聚类6为高校科学数据引用与开放获取，侧重于高等院校层面的科学数据管理与开放获取研究；（7）聚类7为数据管理政策，侧重对数据质量、数据集成、数据可获得性等数据管理元素进行分析。

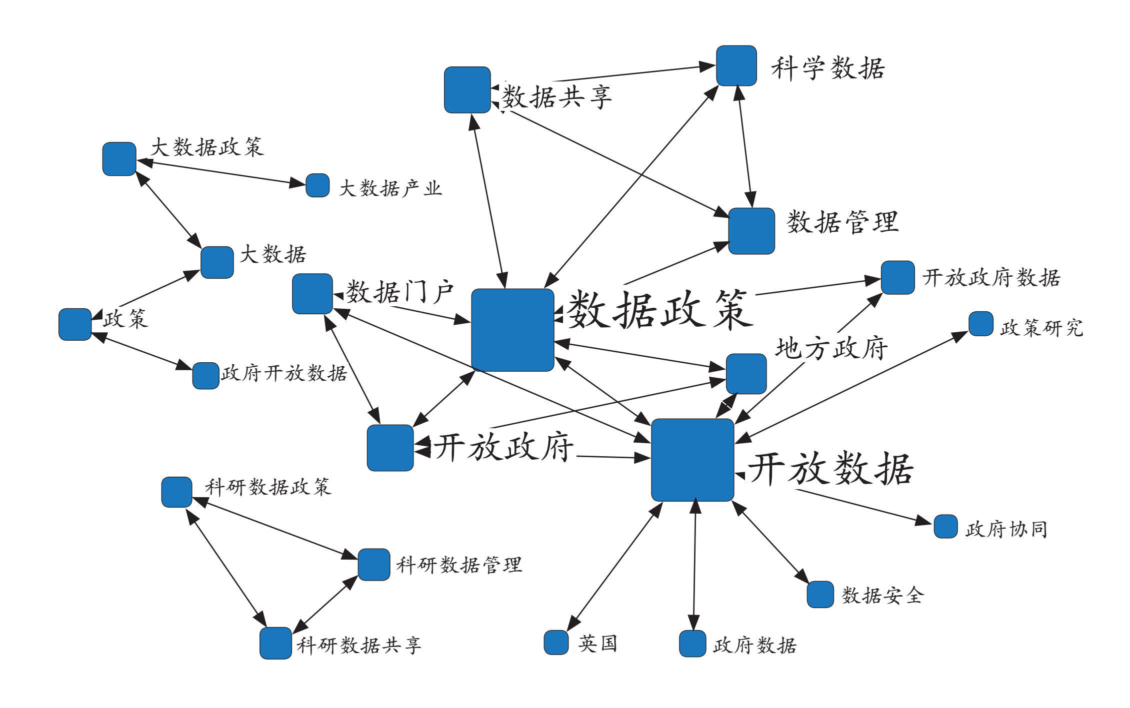


图4　国内科学数据政策研究领域高频关键词分布图谱（词频≥2）

表5　国际科学数据政策研究领域的7个主题聚类

聚类编号	聚类名称	关键词数目	关键词
1	科学数据管理	4	scientific communication, scholarly mmunication, data management, reproducibility
2	科研数据开放与共享原则	6	journal research data policies, FAIR data, scholarly publishing, research data, data sharing, open data
3	科研诚信与科研透明度	3	high impact addiction journals, research transparency, research integrity
4	期刊数据共享	3	resource sharing, journals, collaboration
5	欠发达国家科研数据管理研究	5	Africa, Kenya, ethics, research data management, researchers
6	高校科学数据引用与开放获取	4	science policy, higher education, data citation, open access
7	数据管理政策	6	data quality, data integration, databases, data availability, data policy, policy

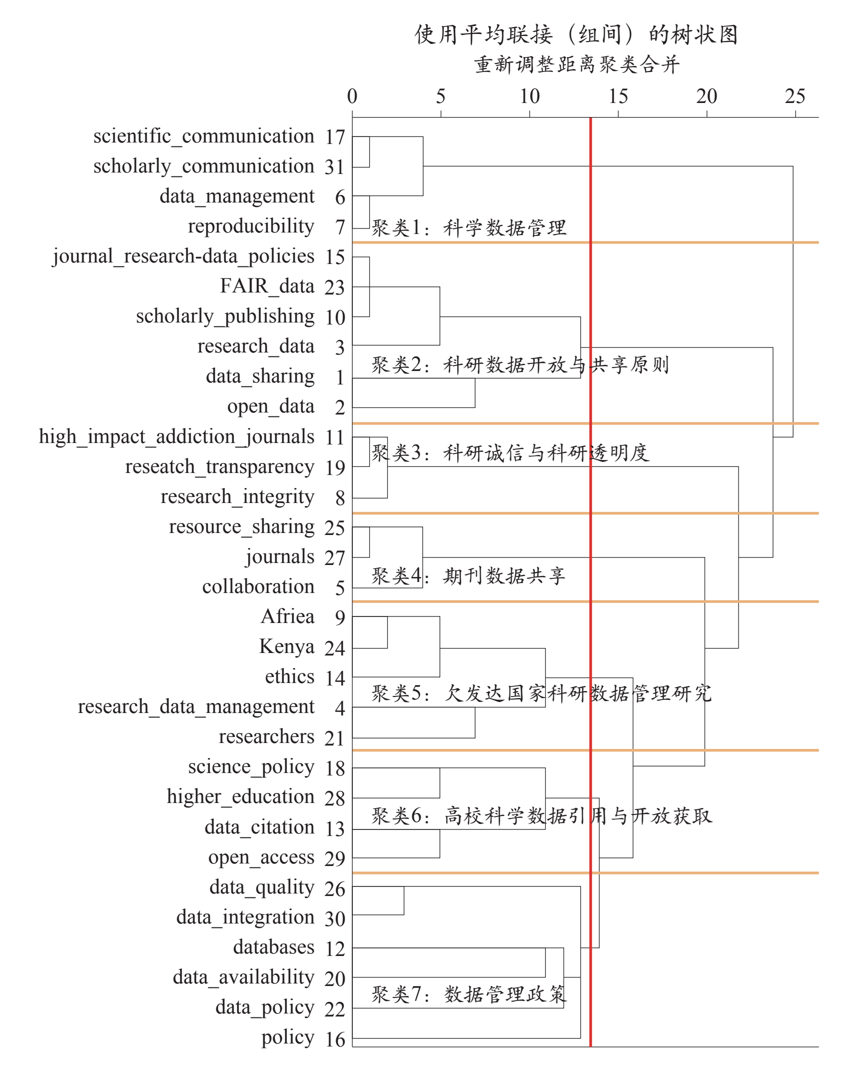


图5 国际科学数据政策研究领域高频关键词系统聚类树状图

3.5.2 国内高频关键词聚类分析

同样对国内科学数据政策研究领域的高频关键词进行系统聚类，得到国内高频关键词系统聚类树状图（参见图6）。

国内高频关键词聚类群体划分情况如表6与图6所示，上述高频关键词可分成13个聚类，分别代表国内科学数据政策研究领域的13个重要研究主题：（1）聚类1、11均为科研资助机构科学数据政策研究，侧重通过内容分析方法分析科研资助机构的科学数据管理、开放及共享政策；（2）聚类2为科研人员关于科研数据保存与共享认知程度的研究，主要从科研数据保存与共享两个角度进行研究；（3）聚类3、4为科研机构科学数据政策研究，侧重于科研机构内科研数据管理与开放政策研究；（4）聚类5、13分别为美国科学数据开放研究与英国、澳大利亚科学数据政策研究，两者均对国际科学数据政策进行统计分析研究；（5）聚类6为高校科研数据管理，主要是分析论述高校科研数据管理的现状与政策；（6）聚类7为高校图书馆科学数据管理，侧重从高校图书馆角度开展科学数据政策研究；（7）聚类8为科研数据管理与共享政策，侧重于科研数据管理和共享方面的研究；（8）聚类9为科学数据开放获取的安全问题研究，主要从数据安全角度分析科学数据开放获取过程中的挑战；（9）聚类10为国际组织科学数据开放共享，侧重从国际组织的角度出发，对科学数据开放共享进行研究；（10）聚类12为学术期刊政策研究，主要通过文献计量方法进行学术期刊政策研究。

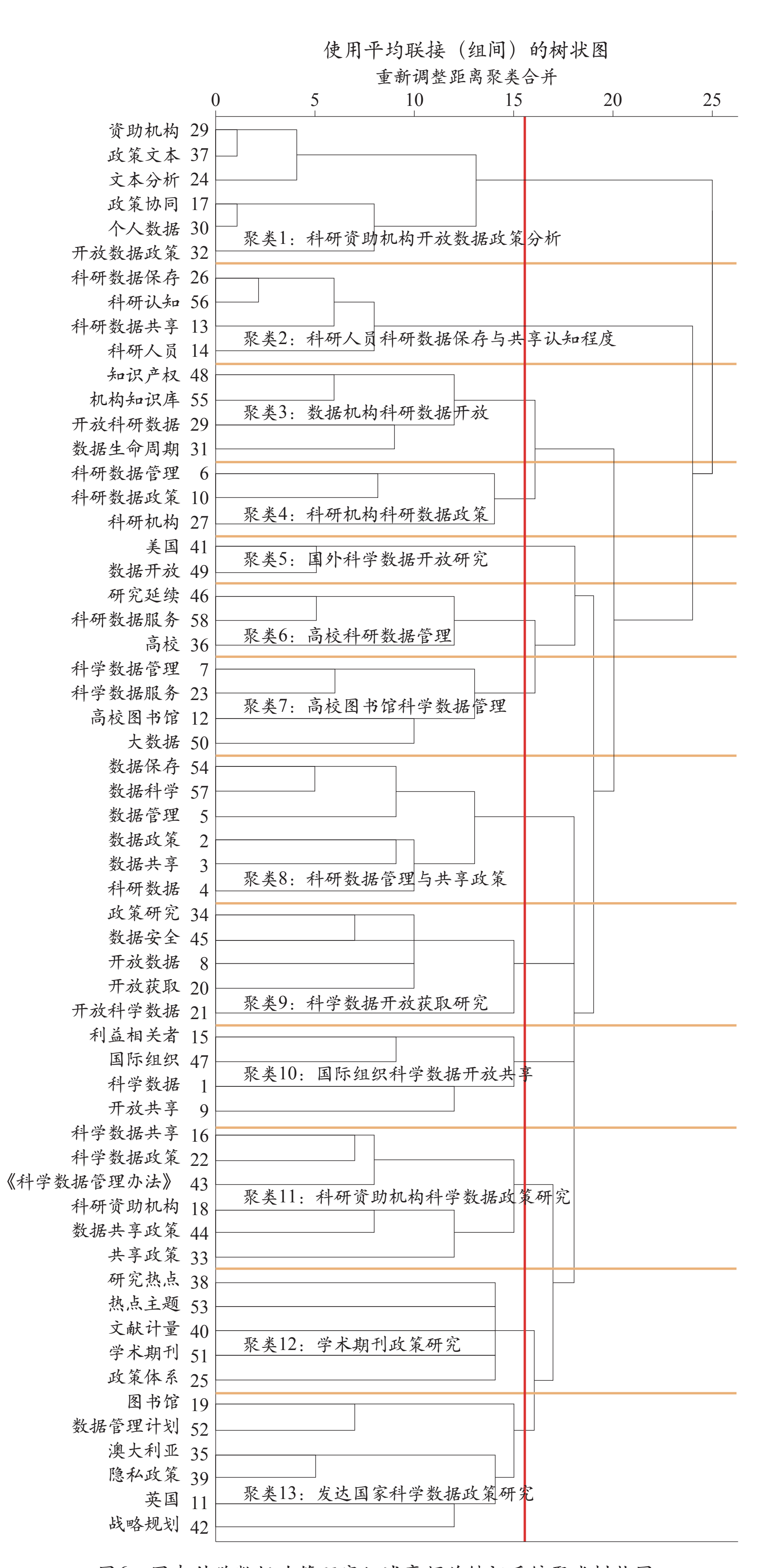


图6　国内科学数据政策研究领域高频关键词系统聚类树状图

表6　国内科学数据政策研究领域的13个主题聚类

聚类编号	聚类名称	关键词数目	关键词
1	科研资助机构开放数据政策分析	6	资助机构、政策文本、文本分析、政策协同、个人数据、开放数据政策
2	科研人员科研数据保存与共享认知程度	4	科研数据保存、科研认知、科研数据共享、科研人员
3	数据机构科研数据开放研究	4	知识产权、机构知识库、开放科研数据、数据生命周期
4	科研机构科研数据政策	3	科研数据管理、科研数据政策、科研机构
5	美国科学数据开放研究	2	美国、数据开放
6	高校科研数据服务	3	高校、研究综述、科研数据服务
7	高校图书馆科学数据管理与服务	4	高校图书馆、科学数据管理、科学数据服务、大数据
8	科研数据管理与共享政策	6	数据保存、数据科学、数据管理、数据政策、数据共享、科研数据
9	科学数据开放获取研究	5	政策研究、数据安全、开放数据、开放获取、开放科学数据
10	国际组织科学数据开放共享	4	国际组织、利益相关者、科学数据、开放共享
11	科研资助机构科学数据政策研究	6	科研资助机构、科学数据共享、科学数据政策、科学数据管理办法、数据共享政策、共享政策
12	学术期刊政策研究	5	研究热点、热点主题、文献计量、学术期刊、政策体系
13	发达国家科学数据政策研究	6	英国、澳大利亚、图书馆、数据管理计划、隐私政策、战略规划

综上，国际、国内关键词群集存在交叉共现部分，主要包括科学数据的管理与共享策略、大数据背景下科学数据政策研究。不同之处在于国际研究侧重于科学数据管理、开放与共享，大量研究聚集在科学数据共享与开放数据政策。国内学者则更多对科学数据政策展开研究。

3.6　内容分析

内容分析法是一种对研究对象内容展开深入分析，进而由表及里、揭示本质内容的科学方法[21]，其目的在于探索研究文献的实质或趋势，已被广泛应用于分析图书情报领域研究主题分布、研究理论应用及研究方法运用[22]。本节采用内容分析法进一步分析国际、国内科学数据政策研究领域所涉及的主题领域。

3.6.1　国际文献研究主题分析

将80篇科学数据政策研究领域外文研究文献涵盖的研究主题分为4个主题类目，并基于各篇文献的具体研究内容，在主题类目下设置二级类目，同时依据每篇文献的具体研究主题，将其划分至对应二级类目（参见表7）。

由表7可知，国际科学数据政策研究主要包括科学数据共享（55%，C类）主题领域，其次是科学数据管理（20.01%，A类）和科学数据政策（16.24%，B类）两个主题领域，再次是开放数据（8.75%，D类）主题领域。

为进一步分析国际科学数据领域研究文献主题的演进趋势，依据文献年度发文量趋势将上述80篇外文科学数据政策相关研究文献按照2007－2010年、2011－2012年、2013－2014年、2015－2016年、2017－2018年、2019－2021年6个时段划分至上文4个主题类目（参见表8）。

据表可知，2007－2010年、2011－2012年与2013－2014年3个时段的相关研究成果较少，涉及“科学数据

表7　国际科学数据政策研究文献主题分析统计表

主题类目	论文数量/篇	百分比/%	二级类目
A.科学数据管理	17	20.01	A1数据长期保存（1）、A2学位论文数据管理（1）、A3 FAIR原则（1）、A4数据质量框架（1）、A5科学数据管理定性调研（1）、A6科研数据管理案例分析（1）、A7科研数据管理框架（5）、A8 EACR原则（1）、A9科研数据管理影响因素（1）、A10科研数据管理馆员（1）、A11科研数据管理知识培训（1）、A12科研数据管理素养（1）、A13科研数据管理意愿（1）
B.科学数据政策	12	16.24	B1科学数据政策需求（1）、B2数据密集型科学（1）、B3数据驱动方法（1）、B4学术期刊数据政策（3）、B5科研数据保护政策/欧盟（EU）数据保护法规草案（1）、B6科研数据保存政策（1）、B7科研数据政策注册中心（1）、B8科研数据开放政策（1）、B9科研数据管理政策调查分析（1）
C.科学数据共享	44	55.00	C1科学数据共享社区（1）、C2科学数据共享策略（3）、C3科学数据共享意愿（7）、C4科学数据集成（1）、C5科学数据共享框架（2）、C6科技期刊数据共享（10）、C7科学数据共享影响因素（3）、C8科学数据共享声明（2）、C9全球科学资料库（1）、C10科研数据隐私保护（1）、C11共享研究数据集（1）、C12科研数据共享标准（4）、C13科研数据共享策略（3）、C14公共卫生研究数据共享（1）、C15生命观察-希腊研究基础设施（1）、C16科研数据二次利用（1）、C17共享科研数据（1）、C18科研数据共享系统（1）
D.开放数据	7	8.75	D1期刊出版商数据开放获取政策（1）、D2开放数据政策（1）、D3开放式科学（1）、D4虚拟研究环境（1）、D5开放获取标准（1）、D6开放科学数据政策（1）、D7开放科研数据（1）

表8　国际科学数据政策研究文献主题时间序列分析统计表（单位：篇）

主 题	2007－2010	2011－2012	2013－2014	2015－2016	2017－2018	2019－2021	总计
科学数据管理	0	0	0	8	2	7	17
科学数据政策	1	1	1	4	2	3	12
科学数据共享	4	4	5	8	5	18	44
开放数据	0	0	1	2	2	2	7
总计	5	5	7	22	11	30	80

政策”“科学数据共享”与“开放数据”3个主题领域。2015－2016年的发文量增速明显，这一时期备受研究人员关注的主题领域是“科学数据管理”与“科学数据共享”，其中“科学数据管理”是该时期新兴的主题领域。2017－2018年时段最受研究关注的主题领域是“科学数据共享”。2019－2021年发文量继续快速增长，除“科学数据共享”外，“科学数据管理”主题领域重新受到更多的研究与关注。

3.6.2　国内文献研究主题分析

按上述方法得到科学数据政策研究领域中文研究文献的研究主题分析表（参见表9）。国内科学数据政策研究的关键主题领域是科学数据政策（46.82%，B类），其次为科学数据管理（27.78%，A类）、科学数据共享（13.49%，C类）、开放数据（7.94%，D类），只有少数研究文献属于数据权益（3.97%，H类）主题领域。

表9　国内科学数据政策研究文献主题分析统计表

主题类目	论文数量/篇	百分比/%	二级类目
A.科学数据管理	35	27.78	A1科学数据管理服务（4）、A2科学数据中心（1）、A3科学数据管理平台（1）、A4科学数据管理计划（1）、A5科学数据管理培训（1）、A6科学数据管理机构（1）、A7科学数据生命周期（3）、A8《科学数据管理办法》（2）、A9科学数据管理服务团队（4）、A10科学数据管理框架（2）、A11科学数据备份与保存（1）、A12科学数据质量审核（1）、A13科学数据管理现状综述（6）、A14数据管理机制（1）、A15数据管理计划（1）、A16科学数据管理要素（1）、A17 E-Science核心计划（1）、A18科研数据管理系统（1）、A19科研数据管理工具（1）、A20科研数据管理调研分析（1）
B.科学数据政策	59	46.82	B1科学数据管理政策（19）、B2科学数据开放获取政策（3）、B3科学数据保存政策（2）、B4科学数据共享政策（18）、B5学术期刊数据政策（6）、B6个人隐私保护政策（5）、B7科研资助机构数据政策（4）、B8出版机构数据政策（1）、B9科研机构数据政策（1）
C.科学数据共享	17	13.49	C1科学数据开放共享（5）、C2国际共享原则（1）、C3共享模式（2）、C4数据授权（1）、C5科学数据共享调查分析（4）、C6 Linked Science制度（1）、C7共享机制（1）、C8 FAIR原则（1）、C9共享科研数据（1）
D.开放数据	10	7.94	D1开放数据政策（3）、D2开放科研数据政策（1）、D3开放科学数据政策（2）、D4开放科学数据认知（1）、D5 FAIR原则（1）、D6开放科研数据（2）
H.数据权益	5	3.97	H1科研数据内容重用（1）、H2科学数据权益保护（1）、H3知识产权（1）、H4权益管理（1）、H5利益相关者（1）

为进一步了解国内科学数据领域研究文献主题的演进趋势，依据文献年度发文量趋势将上述126篇科学数据政策研究领域的中文文献按照2008－2010年、2013－2014年、2015－2016年、2017－2018年、2019－2021年5个时段划分至上文5个主题类目（参见表10）。

据表10可知，2008－2010年时段内相关研究成果较少，仅涵盖“科学数据政策”和“科学数据共享”两个主题领域。2013－2014年、2015－2016年这两个时段的发文量开始呈现快速增长趋势，说明国内学者自2013年起较多关注科学数据政策研究领域，此外，该时期的重点研究主题领域有“科学数据政策”和“科学数据共享”，“科学数据管理”“数据权益”与“开放数据”是该时期的3个新兴研究主题领域。2017－2018年时段内发文量较之前呈现更大幅度的增长，“科学数据政策”是该时期最受关注的主题领域，其次是“科学数据管理”和“开放数据”，此外，“数据权益”经过2015－2016年的空白期后重新受到业内学者的关注。2019－2021年这一时段的发文量继续快速增加，除“科学数据管理”和“科学数据政策”两大老牌主题领域外，“科学数据共享”跻身较受关注的主题领域，开始受到更多学者的关注与研究。

表10　国内科学数据政策研究文献主题时间序列分析统计表（单位：篇）

主 题	2008－2010	2013－2014	2015－2016	2017－2018	2019－2021	总计
科学数据管理	0	1	4	9	21	35
科学数据政策	3	7	13	16	20	59
科学数据共享	3	5	2	1	6	17
开放数据	0	0	2	4	4	10
数据权益	0	2	0	2	1	5
总计	6	15	21	32	52	126

4 结论

互联网与大数据技术的快速发展促使科研工作的成果更多呈现于数字化资源平台，这对科学数据的管理工作提出了更高的要求，合理规范的科学数据政策已逐渐成为科学研究工作向前发展的重要基础性条件。国际、国内科学数据政策研究领域的文献数量呈现波动上升态势；该领域高频发文期刊有*Scientometrics*、*PLoS One*、*Journal of the Association for Information Science and Technology*、《图书情报工作》《情报杂志》《情报理论与实践》等；研究方向以信息科学与图书馆学为主导，但同时也存在部分交叉学科成果；国际该领域研究的高频关键词是："data sharing""open data""research data management""research data"，国内文献的高频关键词是："数据政策""数据管理""科学数据""科学数据管理"。

整体来看，国际、国内科学数据政策研究领域的发展态势有以下几个方面的表现。

（1）在研究成果增长趋势方面，国际、国内的发文量均呈现萌芽增长至波动上升的态势，预测未来将会受到来自更多学科、机构的学者关注，属于发展中的热点研究领域。

（2）在研究主导学科领域方面，该领域研究主要发表在信息科学与图书馆学领域专业期刊，但也存在计算机科学、公共卫生与环境卫生等其他学科专业刊物发表的情况，即国际、国内该领域研究均以信息科学与图书馆学方向为主，同时部分研究成果属于交叉学科范畴。

（3）在研究热点方面，国际、国内科学数据政策研究领域的主要研究热点包括科学数据共享、科学数据管理、开放科学数据等。此外，国际研究侧重于科学数据共享、科学数据管理等主题，国内学者则较多开展科学数据政策方面的研究。总体来看国际学者侧重科学数据的管理与共享、国内学者则更青睐科学数据政策相关研究。

（4）在研究主题演化趋势方面，国际早期研究主题涉及"科学数据政策""科学数据共享"与"开放数据"，2015年后"科学数据管理"与"科学数据共享"成为新的研究主题，开始受到更多的研究与关注。国内早期研究主题较多为"科学数据政策"和"科学数据共享"。自2013年起，国内学者更多地开展"科学数据政策"研究，这一时期的重点研究主题还包括"科学数据共享""科学数据管理""数据权益"与"开放数据"。

参考文献

[1] 王丹丹. 英国《开放科研数据协议》的解读与启示[J]. 数字图书馆坛, 2018, (08): 25-30.

[2] 邢文明, 华小琴. 澳大利亚高校科研数据政策内容分析[J]. 知识管理论坛, 2018, 3(05): 301-312.

[3] 唐义, 张晓蒙, 郑燃. 国际科学数据共享政策法规体系:Linked Science制度基础[J]. 图书情报知识, 2013, (03): 67-73.

[4] 迟玉琢. 科学数据能力研究：内涵、框架和影响因素[J]. 农业图书情报学报, 2020, 32(01): 23-29.

[5] Outline of a Research Data Management Policy for Australian Universities / Institutions [EB/OL]. https://www.rdc-drc.ca/wp-content/uploads/Review-of-Research-Data-Repositories-2015.pdf, [2021-03-15].

[6] UK Research and Innovation. Common Principles on Data Policy [EB/OL]. http://fffg8ed76bbfb4194ad1adad59591f163e1asokn5qnoqvfnu6x5n.fgfy.hlju.cwkeji.cn/funding/information-for-award-holders/data-policy/common-principleson-data-policy/, [2021-03-15].

[7] 张晓青, 盛小平. 国外科学数据开放共享政策述评[J]. 图书馆论坛, 2018, 38(08): 147-154.

[8] 姜恩波, 李娜. 中国开放政府农业数据分析与评价[J]. 农业图书情报学报, 2020, 32(10): 4-15.

[9] 邢文明. 我国科研数据管理与共享政策保障研究[D]. 湖北省武汉市: 武汉大学, 2014.

[10] 中华人民共和国国务院.《科学数据管理办法》[EB/OL]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm, 2021-03-10.

[11] 中国科学院.《中国科学院科学数据管理与开放共享办法（试行）》[EB/OL]. http://www.cas.cn/sygz/201902/t20190221_4679910.shtml, 2021-03-10.

[12] 邢文明, 汤雅静, 秦顺. 国外教育机构科研数据管理政策大纲解读及启示[J]. 数字图书馆论坛, 2019, (05): 9-16.

[13] 姜鑫, 王德庄, 马海群. 关键词词频变化视角下我国"科

学数据”领域研究主题演化分析[J]. 现代情报. 2018, 38(01): 141-146+161.
[14] 张涛，马海群. 国际数据政策文献研究现状、前沿热点与演化趋势分析（1989-2018年）[J]. 情报科学, 2020, 38(09): 149-156.
[15] 孙俐丽，袁勤俭. 数据质量研究述评：比较视角[J]. 农业图书情报, 2019, 31(7): 4-13.
[16] Tancoigne E, Barbier M, Cointet JP. The place of agricultural sciences in the literature on ecosystem services[J]. Ecosystem Services, 2014, 10(S1):35-48.
[17] 柴苗岭，黄琳，任运月. 重要开放农业科学数据资源建设现状综述[J]. 农业图书情报学报, 2020, 32(10): 25-34.
[18] 马海群，姜鑫. 我国档案学研究主题的知识图谱绘制——以共词分析可视化为视角[J]. 档案学研究, 2014, (05): 7-11.
[19] 姜鑫，马海群. 2014年我国图书情报学研究进展述评——基于CSSCI来源期刊的文献计量分析[J]. 现代情报，2016, 36(04): 170-177.
[20] 陈青云，曹建飞，陈荣祯. 从叙词表到知识图谱的构建研究与实践[J]. 农业图书情报, 2019, 31(1): 44-53.
[21] 邱均平，曾倩. 近五年来国内外索引研究的进展与趋势[J]. 情报科学, 2011, 29(07): 961-966.
[22] 姜鑫. 国际图书情报领域“科学数据”研究进展述评——基于SCI/SSCI期刊论文的内容分析[J]. 现代情报, 2018, 38(12): 144-150.

Analysis on Hot Spots and Evolution Trend of Scientific Data Policy Research

Song Yonghui[1,2,3], *Jiang Xin*[2], *Ma Tingcan*[1,3*]
1 Department of Library, Information and Archives Management, University of Chinese Academy of Sciences, Beijing 100190, China
2 College of Information Management, Heilongjiang University, Harbin 150080, China
3 Wuhan Library of Chinese Academy of Science, Wuhan 430071, China
Corresponding Author, E-mail: matc@whlib.ac.cn

[Abstract] At present, scientific data management is being highly concerned and closely studied by scientific research institutions all over the world. Comparing and analyzing the similarities and differences between international and domestic scientific data policy research is conducive to the development of this field. This paper selects the papers related to scientific data policy as the research object, uses the methods of literature metrology and content analysis to analyze the development characteristics and evolution trend of international and domestic research in this field, and reveals the research hot spots. The results show that both international and domestic scientific data policy research are in the stage of exploration and development, and the amount of research literature fluctuates and increases. The research direction is dominated by information science and library science, and some research belongs to interdisciplinary achievements. International research pays more attention to scientific research data sharing, while domestic research focuses on scientific data policy, data management and the research relationship between scientific research data.
[Keywords] scientific data policy, evolution trend, research hotspots, bibliometric, content analysis